Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 97 (2013) 405 – 413

Procedia
Social and Behavioral Sciences

The 9th International Conference on Cognitive Science

Bilateral account of multimodal grounding of meaning

Kawai Chui*

Graduate Institute of Linguistics, No.64, Sec.2, ZhiNan Rd., Wenshan District, Taipei City, 11605, Taiwan

Abstract

This study investigates how the speaker provides and grounds meaning via gestural repetition and speech when participants jointly establish meaning in conversation. It is found that the speaker conveys new meaning with a different linguistic expression, while the addressee's previous gesture for the same reference is mimicked. This multimodal grounding strategy facilitates simultaneous realization of shared knowledge in gesture and new meaning in speech within a clause. It also supports the bilateral process of speaking: The speaker not only provides meaning, but s/he grounds meaning by considering the addressee's knowledge state about the reference. Then, the addressee displays understanding accordingly.

Keywords: mimicked gestures, bilateral process, grounding of meaning, multimodal communication

1. Introduction

Grounding is an important aspect of language use [1-5]. "In dialogue, speakers try to ground their communicative acts as they go along: They work with their partners to reach the mutual belief that the partners have understood them well enough for current purposes" [5: 63]. In Clark and Krych's [5] study, pairs of participants – each pair comprised a Director and a Builder - engaged in the task of assembling ten Lego models. It was found that whether the Director could see the Builder, and whether both were given instructions by audiotape affected the participants' linguistic and gestural performances in grounding the Lego pieces. The findings provide evidence that monitoring the addressees' workspaces is crucial to making grounding more efficient. Speaking is, thus, a bilateral process: "Speakers monitor not just their own actions, but those of their addressees, taking both into account as they speak. Addressees, in turn, try to keep speakers informed of their current state of understanding" [5: 62].

In the process of providing and grounding meaning for mutual understanding of the same reference across turns, speakers can employ multimodal resources to achieve the goal, since "the body-in-action is available as a situated social resource" [6: 250]. Holler and Wilkin's [7] study further demonstrated that when participants in face-to-face dialogues talked about a set of geometrical figures, the speaker would ground a reference by repeating the addressee's previous gesture during the referential communication task. Since everyday conversations provide the most natural sequential context for the examination of multimodal resources, the present study aims to use the conversational data and investigate how the speaker provides and grounds meaning via speech and gestural repetition when the participants jointly establish meaning for the same reference. Furthermore, while grounding can

* Corresponding author. Tel.: +886-02-29393091
E-mail address: kawai@nccu.edu.tw

doi:10.1016/j.sbspro.2013.10.252

provide the clearest evidence for speaking as a bilateral process [5: 63], this study shows that the multimodal way meaning is grounded provides empirical evidence in support of the bilateral claim, manifesting what speakers take addressees into account for and how speakers use that information during the construction of their own linguistic-gestural units, and also how addressees tell speakers about their understanding.

Different from the past task-based studies, the domain of analysis in the naturally occurring conversational data is 'the stretch of talk' different speakers engaged in for the establishment of meaning of the same reference. It also consists of a pair of similar gestures produced by different speakers to depict the same reference across sequential turns. Gestural repetition is 'gestural mimicry' [8: 41] that involves "the recurrence of the same or similar gesture across speakers." Gestures of this kind were also called 'return gestures' [9], 'gestural rephrasings' [10], 'mimicking gestures' [8] or 'mimicked gestures' [7, 11]. In this paper, 'mimicked gesture' is used for gestural repetition.

In the next section, the data for the present study will be introduced. Sections 3 and 4 provide empirical evidence based on the sequential organization of talk in support of the bilateral process of grounding meaning by speech and gestural repetition. Section 5 is the conclusion.

2. Data and methods

The study was based on the data from daily face-to-face conversations in Mandarin Chinese. All the participants were paid, and they were told that they would participate in research on conversation. The stretches of talk for the study met two criteria: First, each stretch involves different speakers holding a discussion about the meaning of a reference. Second, it includes a pair of similar gestures produced by different speakers to depict the same reference sequentially across turns. For the investigation of speech-gesture collaboration in grounding meaning, gestures conveying substantive meaning, i.e., iconic and metaphoric gestures, are considered. For instance, there is a topic about removing a hornet beehive from a tree. The speaker F2 tells that a fireman went up a ladder and then took the beehive down, as shown in Line 1 in Example 1.

(1) 1 F2: ...ta bān le tīzi gùoqù.. ránhoù yí shàngqù jiù zhèyiàng... jiù **zhāixiàlái** le
3SG move PRF ladder go.there then as go up just like this just take off PRT
'He moved a ladder there, and as he went up, like this, (he) just took (it) off.'
2 F1: ..jiù yòng shoǔ bǎ tā **niǎnxiàlái** o
just use hand BA 3SG pluck off PRT
'(He) just used hands to pluck it off?'

A gesture is also produced for the action *zāi* 'pick' *xiàlái* 'down' simultaneously (1a in Fig. 1): F2's right hand rises above head level, with the thumb and index finger extended, and turns clockwise once. This iconic gesture is substantive since it enacts a way the beehive was taken down. In the next turn, this substantive gesture is mimicked by another speaker F1 (1b in Fig. 1) while characterizing the action as *niǎn* 'pluck' *xiàlái* 'down' in the utterance (Line 2 in Example 1). Metaphoric gestures representing abstract ideas also bear a direct semantic relation with utterances, as indicated by the turning-the-hand-in-space gesture for the process of change in Example 2 in Section 3.

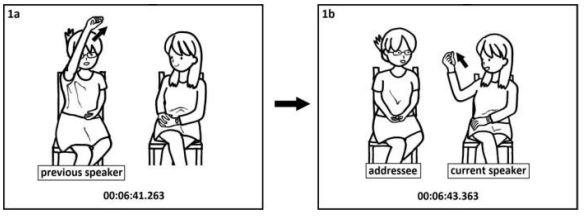

Fig. 1. Gestural depiction of 'pick-beehive-down'

Two coders worked separately to identify and analyze data based on the two criteria. For each eligible stretch of talk, coders judged whether the two similar gestures referred to the same reference based on the information in the sequential context and the content of utterances. A total of twelve instances of mimicked gestures were found. They mostly occurred in the turn right after their first occurrences in the immediately preceding turn. They constituted

twelve pairs of gestures for the analysis of gesture features. Five gesture features were used to evaluate the similarity of gestural forms: 'handedness', 'position', 'orientation', 'hand shape', and 'motion' [12-14]. Coders reached high agreement, 93% on average, with regard to whether the two gestures in each pair shared the same or different features. Total agreement was also reached for the categorization of mimicked gestures: six were iconic; six were metaphoric. These gestures, which are co-referential with their respective counterparts in the prior turns and maintain high similarity across the five gesture features, reveal the way speakers take addressees into account in the bilateral process of providing and grounding meaning during speaking.

3. Contextual situations and the joint establishment of meaning

To understand how speakers ground meaning via speech and gestural repetition for mutual understanding, it is first necessary to discuss the prior contextual situations that readily call for the joint establishment of meaning. Four types of situations in the immediately prior turn were identified: 'verbalization difficulty', 'lack of clarity', 'alignment', and 'disagreement'. In the discussion below, the participant in the prior turn is the 'previous speaker', as distinguished from the 'current speaker' who responds to the call and grounds meaning with multimodal resources. In the conversational data, a previous speaker encounters speaking difficulty when she makes a general statement that if the person whom someone fails to establish a close relationship with, s/he would idealize the person. See Line 1 in Example 2. A gesture is made to depict the verb *lǐxiǎnghuà* 'idealize' (2a in Fig. 2): The right hand rises to cheek level with fingers slightly apart and bent, after which the hand turns around clockwise. After the assertion, the speaker attempts to further explicate what she means by idealization, yet fails to come up with a word after uttering the second degree adverb *hěn* 'very'. The current speaker then resolves the difficulty by providing a different interpretation of idealization later in Line 2.

(2) 1 F1: ..nǐ débúdào de dōngxī.. ránhoù yìzhí dùi tā hěn **lǐxiǎnghuà** de.. **hěn**
2SG NEG.get DE thing then continuously to 3SG very idealize DE very
'For the things that you can't get, then (you) very much keep idealizing him..very'

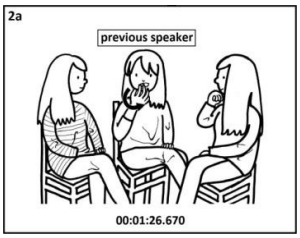

Fig. 2. Gestural depiction of 'idealization'

2 F3: ... nǐ bǎ tā měihuà... dùi
2SG BA 3SG beautify right
'You beautify him.'

In another situation, the need for joint establishment of meaning arises as the meaning of a new reference in the prior turn lacks clarity. It happens when new references are expressed by demonstratives, non-conventional ideophones, or homonyms. In the data, during the assessment of a friend's unusual behaviour, the previous speaker produces an ideophone [yuyu] (Line 1 in Example 3) accompanied by a gesture depicting the strange behaviour (3a in Fig. 3): Speaker's both hands at both sides of the body, with fingers being together and curled into a fist like the ASL 'A' handshape, move slightly up and down alternately. In the absence of explicit meaning in the prior context, the speaker in the next turn (Line 2) needs to provide and ground meaning for the ideophone.

(3) 1 F1: ..xiànzài biànchéng **yuyu** zhèyiàngzhi
now become yuyu like this
'Now, (he) became 'yuyu', like this.'

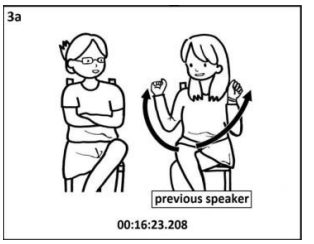

Fig. 3. Gestural depiction of 'unusual behaviour'

2 F2: ..ránhoù hái hùi tiàowǔ
then also will dance
'(He) will also dance.'

The third situation has to do with alignment, in that what the previous speaker talks about is recognized by the current speaker. The data includes a conversational topic about feeling itchy during the collection of grains in the field. The previous speaker M3 explains that there are prickles *máng* on the grains (Line 1 in Example 4). At the same time, a gesture is also produced for prickles (4a in Fig. 4): Speaker's right hand rises to shoulder level, with fingers together and curved into the palm; his left hand goes to chest level, with fingers together. The configuration as a whole enacts holding the stem of a grain on which there are prickles. The current speaker accepts the idea by providing more information in the next turn (Line 2).

(4) 1 M3: .. yīnwēi tā yoǔ [**máng** a]
because 3SG have prickle PRT
'Because it has prickles.'

Fig. 4. Gestural depiction of 'prickles'

2 M1: ..shàngmiàn yoǔ nàge háomáo
on there.be that fine hair
'There is fine hair on (it).'

Finally, contrary to the third situation, a current speaker engages in the joint establishment of meaning because s/he does not agree to the content in the prior turn. For instance, in the topic about the kind of musical instrument a character plays in a movie, the previous speaker uses a general term *yuèqí* 'musical instrument' in speech (Line 1 in Example 5) but gestures a particular kind that requires a bow (5a in Fig. 5): Speaker's right hand goes up to should level at her right hand with fingers being curled into a fist as if holding a bow; the left hand rises to waist level, also with fingers being curled into a fist as if holding the lower part of the instrument. Then, the right hand moves to the left one time to enact the idea of playing a stringed musical instrument that requires a bow. Since the current speaker holds a contrary opinion, he thus brings up a different interpretation for *yuèqí* in his turn (Line 2).

(5) 1 F: .. màobó shì nà zǒng... zhūanyiè de nà zǒng... **yùeqì** de.. nǐ
Maobo COP that kind professional DE that kind musical instrument DE 2SG
zhīdào ma.. sǔoyǐ
know PRT so
'Maobo (used) that kind of professional musical instrument, you know. So...'

Fig. 5. Gestural depiction of the kind of musical instrument requiring a bow

2 M: (0) méiyiǔ.. tā shì nàge...(1.1)[<L5 jù xúanzǎi L5>
no 3SG COP that play plucked lute with a wooden body
'No, he played the kind of plucked lute with a wooden body.'

In brief, the four types of contextual situations mentioned above are the previous speakers' turns, which comprise mainly assessments and assertions, and new references are represented in both speech and gesture. Since other speakers have their own understanding, the joint establishment of meaning for the gestural references starts in the next turn.

4. Bilateral account of multimodal grounding of meaning

In the next turn, to express his/her own understanding of a reference being introduced in the prior turn, the current speaker can use multimodal resources to provide and ground the understanding. Clark and Schaefer [15] proposed a pattern of grounding in dialogue: a *presentation phase* followed by an *acceptance phase*. This section investigates the use of language and gesture vis-à-vis the two phases.

In the presentation phase, how does the current speaker provide and ground his/her understanding in each of the four contextual situations? Generally, "the speaker must design what she says against the current common ground with her partner. His beliefs about their common ground should be coordinated with hers if they are to understand one another efficiently" [16:184]. The design is multimodal in face-to-face conversational interaction. First, to cope with the situation of verbalization difficulty, as exemplified by the previous speaker's failure to come up with a word to elucidate *lǐxiǎnghuà* 'idealize' in Example 2, the current speaker designs her utterance with a new statement comprising a different lexical verb *měihuà* 'beautify' (Line 2 in Example 2). In gesture, instead of depicting *měihuà*, simultaneous with the verb is a metaphoric gesture mimicking the first gestural occurrence for *lǐxiǎnghuà*. See 2b in Fig. 6.

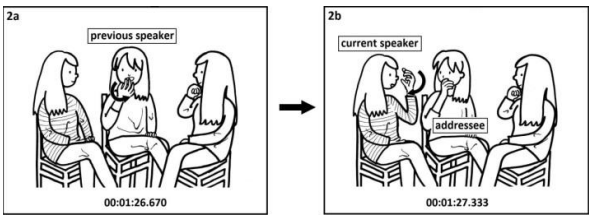

Fig. 6. Gestural repetition of 'idealization'

In the situation of lack of clarity, like the occurrence of the non-conventional idiophone [yuyu] in the data, the current speaker repeats the unusual-behaviour gesture (3b in Fig. 7) at the time she offers explicit meaning for the ideophone by using the new lexical verb *tiàowǔ* 'dance' in her own assessment of the friend (Line 2 in Example 3).

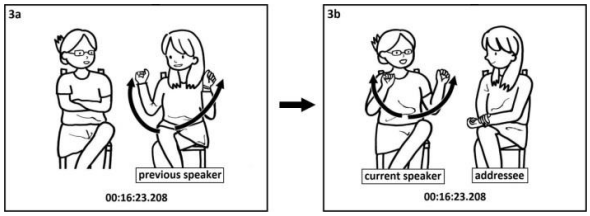

Fig. 7. Gestural repetition of unusual behaviour

Thirdly, to show alignment with the previous utterance, as in the discussion about feeling itchy during the collection of grains, the current speaker mimics the previous speaker's hold-grain-stem gesture (4b in Fig. 8) while characterizing another quality of the prickles - having *háomáo* 'fine hair' on the grains (Line 2 in Example 4).

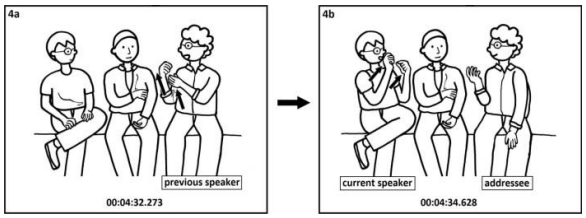

Fig. 8. Gestural repetition of 'prickles'

Finally, to express a contrary opinion, the current speaker can also take the same multimodal design - produce a mimicked gesture while establishing different meaning for the same gestural reference. In the case about musical instrument, the current speaker disagrees with the previous speaker that it is the type that requires a bow. He proposes another type that is played with fingers, as represented in speech by *yuèqìn* 'plucked lute with a wooden body' (Line 2 in Example 5). But intriguingly in gesture, the speaker still produces the same play-with-a-bow gesture (5b in Fig. 9).

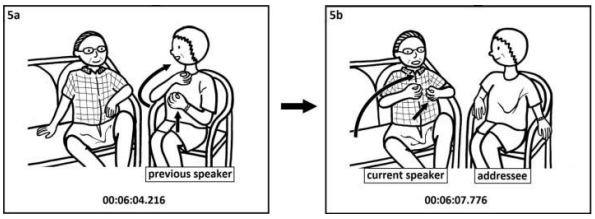

Fig. 9. Gestural repetition of 'musical instrument requiring a bow'

In short, when the current speaker engages in the joint action to provide his/her own interpretation of the gestural reference being brought up in the prior turn, the way the new interpretation is provided and grounded is similar through the collaboration between speech and gesture: During the construction of utterance, a different linguistic expression is used to convey new meaning for the reference under discussion; at the same time, the current speaker mimics the previous speaker's gesture without encoding new information. The result bears out the fact that "participants [in dialogue] work together in determining the course of each utterance. They rely not only on each other's vocal signals, but on each other's gestural signals" [5: 79]. There can be various ways speakers can use to incorporate gestural signals in the process of grounding. According to Clark and Krych [5: 79], participants in task-

based communication relied on gestural signals like "exhibiting, poising, pointing at, and placing physical objects, nodding and shaking heads, and directing eye gaze, and on other mutually visible events." The present study, based on everyday conversational data, found another multimodal grounding strategy – presentation of new reference and simultaneous repetition of the addressee's previous gesture.

In this division of labor between the two modalities, it is intriguing that the current speaker would rather repeat the previous speaker's gesture than produce a different one for the new constituent, even though the current speaker disagrees on the meaning or referent being mentioned in the prior turn. Given the *principle of least joint effort* [2-4], repeating the addressee's previous gesture could be part of the speaker's design, in that providing new meaning on a shared foundation is an efficient way to ground meaning. Using multimodal resources then facilitates the online realization of expressing both shared knowledge in gesture and new meaning in speech at the same time within a clausal unit. Such multimodal grounding strategy supports the bilateral process of speaking, in that the speaker not only provides meaning for the reference at issue, s/he also grounds the new information by taking the addressee into account at the same time. The findings show that the speaker considers the addressee's knowledge state about the same reference. Since the initial occurrence of the mimicked gesture has been produced by the addressee in his/her prior turn, for the speaker to repeat the gesture thus constitutes a semantic foundation of knowledge state shared by the speaker and the addressee, based on which further meaning of the same reference is conveyed in speech.

Finally, as claimed by Clark and Krych [5: 62] about the bilateral process of speaking, "[a]ddressees, in turn, try to keep speakers informed of their current state of understanding." In the conversational data, after the current speaker has presented new meaning, the addressee also displays his/her understanding accordingly. In some cases, the addressee shows disagreement, so that the co-establishment of meaning continues and takes more turns. The acceptance phase comes when confirmation of understanding is expressed. In Mandarin, particles such as *duì* 'right' or *o* 'I see' to convey agreement with the speaker's linguistic-gestural construction, or head nods are frequently used, which can be accompanied by more elaboration to further indicate acceptance. For the talk to go on in the lack of the addressee's explicit display of understanding implicates agreement with no objection. All the various types of responses on the part of the addressees, again, support the bilateral process of speaking, as "addressees take an active part both: (1) by telling speakers about their understanding and (2) by giving them access to evidence of understanding" [5: 77].

5. Conclusion

In daily conversation, the use of mimicked gestures is not frequent. One plausible reason is that speakers perform many other actions besides establishing shared understanding of meaning. Another reason is that speakers do not necessarily repeat others' gestures for grounding. Nevertheless, the use of mimicked gestures is by no means a matter of chance. In the task-based dialogues in Holler and Wilkin [7], the study required the two participants to focus their talk on references, "in order to figure out whether they are talking about the same thing" (ibid: 136). A total amount of 113 mimicked gestures were found. While mimicked gestures in speech communication are not produced by chance, the present study has shown that their occurrence and collaboration with speech in the joint establishment of meaning indicates a multimodal strategy for grounding of meaning. Such grounding strategy, in turn, bears out the bilateral nature of speaking manifesting that speakers take into account addressees' knowledge state to form a semantic foundation across speakers during the construction of their own linguistic-gestural units. The addressees in the next turns also inform the speakers about their understanding. Whether this multimodal grounding strategy being used by Mandarin speakers in conversation is language-specific or not awaits future studies across different languages.

References

[1] Clark HH, Wilkes-Gibbs D. Referring as a collaborative process. *Cognition* 1986; 22:1–39.

[2] Clark HH, Schaefer EF. Concealing one's meaning from overhearers. *J Mem Lang* 1987; 26:209–25.

[3] Clark HH, Brennan SE. Grounding in communication. In: Resnick LB, Levine JM, Teasley SD, editors. *Perspectives on socially shared cognition*, Washington, DC: American Psychological Association; 1991, p. 127–49.

[4] Clark HH. *Using language*. Cambridge: Cambridge University Press; 1996.

[5] Clark HH. H., Krych MA. (2004). Speaking while monitoring addressees for understanding. *J Mem Lang* 2004; 50:62-81.

[6] Lerner Gene H. Turn-sharing. In: Ford CE, Barbara AF, Thompson SA, editors. *The language of turn and sequence*, Oxford: Oxford University Press; 2002, p. 225–56.
[7] Holler J, Wilkin K. Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *J Nonverbal Behav* 2011; 35:133–53.
[8] Kimbara I. On gestural mimicry. *Gesture* 2006; 6:39–61.
[9] de Fornel M. The return gesture: Some remarks on context, inference, and iconic gesture. In: P. Auer P, Luzio AD, editors. The contexualisation of language. Amsterdam: John Benjamins; 1992, p. 159-76).
[10] Tabensky A. Gesture and speech rephrasings in conversation. *Gesture* 2001; 1:213–35.
[11] Parrill F, Kimbara I. Seeing and hearing double: The influence of mimicry in speech and gesture and observers. *J Nonverbal Behav* 2006; 30:157–66.
[12] McNeill D. *Hand and mind – What gestures reveal about thought*. Chicago: The University of Chicago Press; 1992.
[13] McNeill D. *Gesture and thought*. Chicago: University of Chicago Press; 2005.
[14] Kimbara I. Gesture form convergence in joint description. *J Nonverbal Behav* 2008; 32:123–31.
[15] Clark HH, Schaefer EF. Contributing to discourse. *Cognitive Science: A Multidisciplinary Journal* 1989; 13:259–94.
[16] Wilkes-Gibbs D, Clark HH. Coordinating beliefs in conversation. *J Mem Cogn* 1992; 31:183-94.

Appendix 1: Speech transcription conventions

[ ]	speech overlap
...(N)	long pause
...	medium pause
..	short pause
<L5 L5>	switch from Mandarin to Southern Min

Appendix 2: Abbreviations of linguistic terms

2SG	second person singular
3SG	third person singular
BA	morpheme *bǎ*
COP	copula verb
DE	morpheme *de*
NEG	negative morpheme
PRF	perfective morpheme
PRT	discourse particle

Appendix 3: First occurrences and mimicked occurrences of gesture

In each pair of figures, the frame on the left is the first occurrence produced by a 'previous speaker'. The frame on the right is the mimicked occurrence produced by a 'current speaker'; the 'previous speaker' then becomes the 'addressee'.

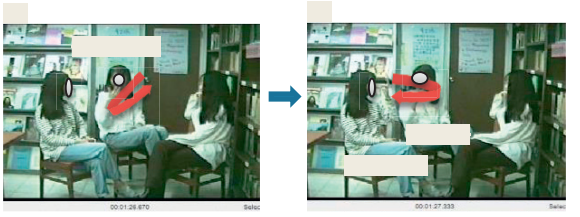

Fig. 10. Gestural repetition of 'idealization'

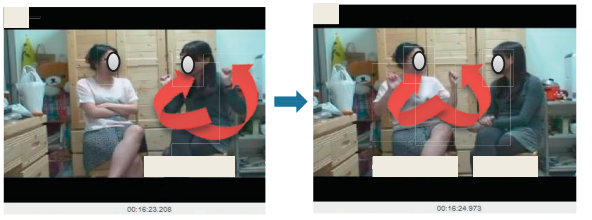

Fig. 11. Gestural repetition of 'unusual behaviour'

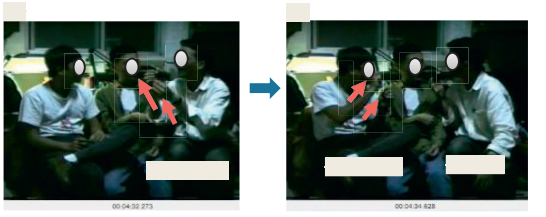

Fig. 12. Gestural repetition of 'prickles'

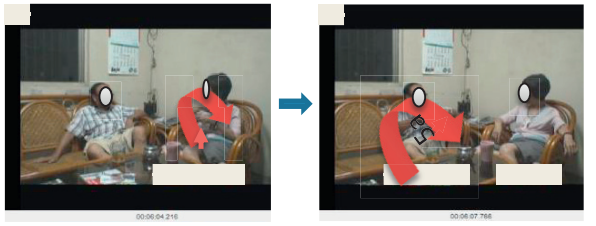

Fig. 13. Gestural repetition of 'musical instrument requiring a bow'